DOCUMENT RESUME

ED 078 075 TM 002 894

AUTHOR Lewis, Michael
TITLE Infant Intelligence Tests: Their Use and Misuse.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RB-73-10
PUB DATE Jan 73
NOTE 19p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Child Development; Cognitive Development; *Infants; Intelligence; *Intelligence Tests; Intervention; Learning Experience; Literature Reviews; *Models; Predictive Ability (Testing); Test Results

ABSTRACT

Data from a variety of infant intelligence scores make clear that it is not possible to consider (1) that infant intelligence is a measurable, stable and unitary construct, (2) that there is a general g factor easily discernible in infancy, (3) that there is stability of scores both within and across scales, or (4) that there is predictability across age. These facts are discussed for their implications for models of intelligence, the use of intelligence tests in infancy, and finally intervention programs. It is concluded that the implicity model of general intelligence rests upon its function for society rather than its scientific merit. An alternative model of infant development is offered which is related to the acquisition of specific skills, the learning of which is dependent upon the match between the subject and the nature of the learning experience. (Author)

RB-73-10

RESEARCH BULLETIN

INFANT INTELLIGENCE TESTS: THEIR USE AND MISUSE

Michael Lewis

This paper has been accepted for publication by Human Development. Any citations should be made to the Journal rather than to the Bulletin.

Educational Testing Service
Princeton, New Jersey
January 1973

INFANT INTELLIGENCE TESTS: THEL. USE AND MISUSE

Michael Lewis

Educational Testing Service

Abstract

Data from a variety of infant intelligence test scores make clear that it is not possible to consider (1) that infant intelligence is a measurable, stable and unitary construct, (2) that there is a general g factor easily discernible in infancy, (3) that there is stability of scores both within and across scales, or (4) that there is predictability across age. These facts are discussed for their implications for models of intelligence, the use of intelligence tests in infancy, and finally intervention programs. It is concluded that the implicit model of general intelligence rests upon its function for society rather than its scientific merit. An alternative model of infant development is offered which is related to the acquisition of specific skills, the learning of which is dependent upon the match between the subject and the nature of the lea~ning experience.

Key words: Infancy, Intelligence, Subject-Treatment Interaction, Intervention

INFANT INTELLIGENCE TESTS: THEIR USE AND MISUSE[1]

Michael Lewis

Educational Testing Service

The concept of intelligence, the belief that it is relatively easily measurable, and that, as a monolithic construct, it is a useful predictor of subsequent human behavior, is firmly engrained in the mind of Western man. Consequently, discussion of this construct has been rendered difficult.

Using the data from a wide variety of infant and young children's tests of intelligence, we shall attempt to review the support for this construct. We will first demonstrate that infant intelligence as a measurable, stable and unitary construct is without foundation in fact. There is no general g factor easily discernible in infancy, no stability of scores both within and across scales, and no predictability. The only way to understand why this information, which in part has previously been known, has gone unused is to observe both the uses and function of IQ scores in a technological society.

The overall theme of this discussion rests upon two points: there is no demonstrable construct as infant intelligence, and early intervention procedures have failed adequately to test their effectiveness because they have neither taken the measurement issues nor the subject-treatment interaction into consideration. From these points of view the implicit model of general intelligence rests upon its function for society rather than its scientific merit. An alternative model of infant development is offered which is related to the acquisition of specific skills, the learning of which is dependent upon the match between the subject and the nature of the learning experience.

Is Infant Intelligence a Measurable, Stable and Unitary Construct?

In common with many others, Burt, Jones, Miller, and Moodie (1934) expressed a view of intelligence as a finite potential with which the individual was endowed at conception, the manifestations of which increased at a stable rate during the growth process but which was subject neither to qualitative change nor to environmental influence. ". . . it is inherited or at least innate, not due to teaching or training. It is intellectual, not emotional or moral and remain. uninfluenced by industry or zeal." Moreover, Burt held that intelligence could be measured with accuracy and ease. It is a sine qua non of such a view that measures of intelligence have high predictive validity from one age to another. Such validity is singularly lacking from every instrument used to assess intelligence during infancy. Bayley (1933) employing the first version of her infant developmental scales reported very little correlation between scores at 1, 2 and 3 months and scores at 18 to 36 months. These correlations range between -.04 and .09.

McCall, Hogarty, and Hurlburt (1972) observed the stability and growth of intelligence in a sample of infants seen in the Fels Longitudinal Sample. Correlations of the Gesell scores in infancy were compared with childhood Stanford-Binet full-scale scores. The results, both for boys and girls, demonstrated that there was relatively little correlational relationship between the Gesell tests and the Stanford-Binet scores. McCall et al. went on to construct a correlational matrix using data from a variety of different investigations. The results compare the relationship of IQ on a variety of infant tests with a variety of childhood tests. The data for over eight different reported studies reveal that there is almost no relationship between the first 12 months of life and subsequent test performance (the highest correlation accounts for just about 10% of the variance). The highest correlation, .54, accounts for less than 30% of the variance between any two ages. The authors conclude that in the first three years of life there is relatively poor prediction in infant tests of intelligence to IQ scores assessed in middle or late childhood. McCall et al. took great pains to find individual or factor item stability across tests and age; nevertheless, they were forced to conclude that even with this type of analysis and the use of a variety of other multivariate techniques, the correlational relationship between different ages "remains modest and of minimum practical utility." In conclusion, they reject the simple conceptualization of a g factor in infancy. "The search for correlational stability across vastly different ages implies a faith in a developmentally constant, general conception of intelligence that presumably governs an enormous variety of mental activities. Under that assumption, the nature of the behavioral manifestations of g would change from age to age, but g itself is presumed constant, and this mental precocity at one age should predict mental precocity at another. Confronted with the evidence reviewed above, this g model of mental development must be questioned (McCall et al., 1972, p. 736)."

Perhaps if we turn from standardized kinds of tests, such as the Bayley or Gesell, to more recent approaches suggested by the Geneva school, we could find stability of infant mental ability. It may be necessary to utilize Piagetian theory and explore tasks more closely related to sensorimotor development to find stability and consistency. King and Seegmiller (1971) applied the Hunt and Uzgiris Scales of Perceptual Cognitive Development (1966) to 14-, 18-, and 24-month-old infants. This test for the measurement of sensorimotor development consists of seven scales. The consistency of scores on these seven scales was compared across three ages as was the relationship at 14 months across the different scales. Not only did the authors find relatively little consistency in terms of the correlational scores (only four out of 24 possible correlations were significant), but also, relatively little consistency across the various scales. Thus, even when we consider the nonstandard intelligence tests and look at sensorimotor development, at least as measured by the Hunt and Uzgiris scales, we find no evidence for a consistency across age nor a g factor.

Lewis and McGurk (1972) obtained and related three different types of infant intelligence tests. Infants were seen longitudinally from three to 24 months at which time they received the Bayley Scales of Infant Development (1969) and the Object Permanence Scale from the Escalona and Corman Sensorimotor Scales (1967). In addition, at 24 months the children received the modified Peabody Picture Vocabulary Test in which both a comprehension and production language score were obtained. For the Bayley Scales the interage correlations proved to be relatively weak. Only two were significant and both of them accounted for less than 30% of the variance. The same was true for the Object Permanence Scales of Sensorimotor Development. Out of a possible 15 correlations across the three to 24 months only two were significant and they accounted for less than 25% of the variance. As in the Bayley scores, the infant's performance on a sensorimotor function followed no clear pattern across age. Lewis and McGurk then observed the correlations between the Bayley and Object Permanence Scales at each age and between language development at 24 months and the Bayley and Object Permanence scores at each age. The results indicated an interesting developmental pattern. First, the Bayley Scales were most closely related to the Object Permanence Scales of the sensorimotor task in the first six months of life, while the Bayley Scales were most closely related to language at 18 and 24 months. This result makes good sense. The early items from the Bayley Scales are closely related to sensorimotor function, while the later Bayley items are related to language. And finally, and most important for our discussion, there was no significant relationship between the Object Permanence Scales of sensorimotor functioning at any age and language ability at 24 months.

Three recent papers (King & Seegmiller, 1971; Lewis & McGurk, 1972; McCall et al., 1972) seem both to extend and reinforce earlier findings and support several broad conclusions concerning intellectual function during infancy: (1) Within a wide variety of standardized tests such as the Bayley and Gesell there is relatively little interage consistency in test performance during the first two years of life. Thus, children who are precocious at one age are not necessarily precocious at another. Moreover, early precociousness in the first two years seems to be unrelated to childhood performance on standard IQ tasks. (2) Nonstandardized tests, constructed out of a Piagetian framework of sensorimotor development, also fail to show any consistency within the first two years of life. Thus, high scores on Object Permanence at one age do not necessarily mean that the child will have high scores on the Object Permanence at other ages. (3) Even within a particular age the results of both King and Seegmiller (1971) and Lewis and McGurk (1972) fail to indicate consistency across different measures of intellectual functioning; for example, there is little relationship between the Bayley Scales and Object Permanence Scales. Moreover, within age there is no consistency for tests such as the various sensorimotor scales of Hunt and Uzgiris (King & Seegmiller, 1971) or across different factors such as those found by McCall et al. (1972) for the Gesell scales.

These results, as well as those reviewed by Thomas (1970), Stott and Ball (1965), and Bayley (1970), support the position that there is no consistency across or within age in a wide variety of tests purported to measure infant mental functioning. Therefore, the conception of a developmentally constant general intelligence is not a very tenable hypothesis.

Uses of Infant IQ Scores

What do these conclusions imply for the notion of intelligence which has been argued to be "inherited or at least innate, not due to teaching or training . . . and remains uninfluenced by industry or zeal [Burt, 1934]"? Such a model of human capacity must clearly be dealt a severe blow from a review of the infancy literature. And yet, such a conception of man remains. While these intelligence scales have thus been acknowledged to have limited functions, they are still widely used in clinical settings in the belief that, although la [illegible] in predictive validity, they provide a valuable aid in assessing the [illegible] health and developmental status of babies at the particular time of testing. This procedure is justified only if in the interpretation of such scores they are regarded solely as measures of present performance and not as indices of future potential. What this performance may mean is questionable since it is possible that superior performance may be indicative of subsequent poor performance. For example, Bayley shows a negative correlation of -.30 between males early in test behavior and IQ at 16 to 18 years (Bayley, 1965). Thus, infant scales are quite invalid as measures of future potential, and it is also unlikely that they properly assess a child's current performance vis à vis other children.

Currently intelligence test scores are widely used as the criterion measure in the evaluation of infant intervention or enrichment programs.

The experimental subjects are compared to the control subjects in terms of their performance on intelligence tests. If the scores of the experimental group are higher than those of the control, the program is evaluated positively; if not, it is evaluated negatively. Implicitly assumed is that infant intelligence is a general unitary capacity and that mental development can be enhanced as a result of the enrichment experience in a few specific areas. Similarly, it is assumed that infant scales are adequate to reflect any improvement that occurs in competence as a consequence of a specific enrichment experience. However, infant intelligence as a general unitary capacity is highly questionable. Moreover, that infant scales are adequate to reflect improvement in specific enrichment experiences must also be highly questioned since both across and within consistency of a variety of infant skills tested show relatively little consistency.

Thus, the data on infant intelligence tests also cast doubt on whether the scores have any generalizability beyond the particular set of abilities or factors sampled at the time of testing. An infant who showed dramatic gains in tasks involving sensorimotor functions would not necessarily manifest such gains in tests involving verbal skills. The implications of these conclusions for a wide variety of evaluative policies concerning infant intervention must be considered. For example, infant intelligence scales, no matter how measured, are quite unsuitable instruments for assessing the effects of specific intervention procedures, primarily because infant intelligence cannot be considered a general unitary trait but is rather a composite of skills and abilities which do not necessarily covary. Such a view of intelligence is by no means new (see, for example, Guilford, 1959), but it is one which must be repeatedly stated in order to counteract the tendency to utilize simple and single measures of infant intelligence. An example will clarify this issue.

Consider an intervention procedure primarily intended to influence sensorimotor intelligence, for example, the development of object permanence. An appropriate curriculum might involve training infants in a variety of peek-a-boo and hide-and-seek tasks. According to the data presented, a standard infant intelligence scale would be the wrong instrument to use in assessing the efficiency of such a program and is likely to lead to erroneous conclusions concerning the program's efficiency. Even more serious is the possibility that by using the wrong instrument of evaluation over a large number of programs one would erroneously conclude that intervention in general is ineffective in improving intellectual ability, thus supporting the genetic bias that environment is ineffective in modifying intelligence. There are few who would suggest that school children should be administered a standard intelligence test after a course in geography, yet such a procedure would be exactly analogous to using an intelligence test to measure the success of teaching the object concept to the young infant. The success of a geography course is best assessed by testing geographical knowledge and understanding and by the same token the success of a program stressing sensorimotor skills is best assessed by specific tests of sensorimotor ability. In both cases there may in some instances be improvement in intelligence test scores but such improvement has to be regarded as fortuitous.

The Function of IQ Scores

Burt's (1934) view cannot be supported by the data. Why then should this view of intelligence hold such a dominant position in the thinking of contemporary scientists and public alike? The answer to such a question may be found by considering the function or use of the IQ score in a technological society. The function of the IQ score is and has always been to help maintain a social hierarchy, the function of which is to create a division of labor within the culture. That is, to determine who will go to school in the first place, who will get into academic programs that lead to college, etc. These divisions in turn determine the nature of labor the child will perform as an adult. This division of labor, a necessity in a complex society, is then justified by scores on a test designed to produce just such a division. If we cannot make the claim that IQ differences at least in infancy are genetically determined, then we must base them on differences in cultural learning.

But these differences, for the sake of the division of labor, are exactly what the IQ tests are intended to produce. The hierarchy of labor is maintained by the genetic myth. The hierarchy produces the test differences and the test differences are used to maintain the hierarchy. Thus, IQ scores have come to replace the caste systems or feudal systems which previously had the function of stratifying society. Wherein these latter systems were supported by evoking the Almighty, the present system evokes Mother Nature. Undoubtedly, some sort of division of labor is necessary. We must find alternative means of achieving it.

An Interactionist Approach

There are, of course, alternative views to Burt's (1934) genetic position. In the present discussion and for the sake of increasing the range of considerations we shall take a totally interactionist view--namely, that experience is both necessary for and the material of knowledge (see Lewis & Lee-Painter, 1972). Like most interactionists, we hold that intelligence (cognitive structures) is the consequence of action in the world and it is influenced by experience.

Cognitive structures are a consequence of interaction for adaptation and it is reasonable to suggest that they are influenced by the nature of the interaction itself. This kind of theorizing suggests an explanation for several different and divergent phenomena. First, it may help to explain individual and cultural differences in thought processes (Cole, Gay, Glick, & Sharp, 1971). Second, it may explain why certain kinds of structures or groupings of structures are no longer capable of maintaining equilibrium. That is, if the world in which assimilation and accommodation take place changes, then the old structures or groupings are no longer adaptive in dealing with what is presently occurring. Thus, rather than emphasizing the genetic underpinning as the pressure for consistent change, we evoke the consistent pressures of the world. Although these are not specifiable at the moment neither are the genetic substrates which evoke the consistency both across the developmental sequence and people. We chose then to avoid relying on a nativist approach (some type of prewiring of sequence) and instead argue for an environmental organismic interaction in the process of development.

The effect of the infant's environment may make even more of a difference when we consider other cognitive structures, those which do not fit under a logical-mathematical framework, that is, space, volume, time, etc., as well as noncognitive structures. The infant and young child certainly develop structures about their social world through assimilation and accommodation. Unlike the logical-mathematical dimensions the specific attributes of the social world are as yet undefined, but there is reason to believe that these structures are affected by what the infant assimilates and accommodates to. For example, each time one infant vocalizes its mother vocalizes back, while for another infant vocalization produces a smile or look. For these two infants their vocalizations produce two worlds--one of vocalization, one of smiling. What is the effect, if any, on the child's resulting cognitive and noncognitive structures? Can we maintain that the resulting structure, infant action-outcome, will be invariant to either condition? Is it possible that in both examples the infant develops the knowledge of its mother (through responsivity toward him), but the nature of that knowledge is related to the specific behaviors directed toward it? We believe that the structures (intelligence), both in the cognitive and in the socioemotional realms of knowing must be affected by the environment in which the structures are formed.

The implications of such an interactionist position can be seen most clearly in how we might organize our intervention programs--that is, change the environment to effect intellectual changes in the infant, and how to measure these changes (Lewis, 1972).

Implications for Intervention Programs

It has been argued that the success or failure of intervention programs in early childhood and infancy is an indicator of the effect of environment on the intellectual growth and capacity of the child. If a variety of intervention programs are shown to be ineffective, then intervention or environment *per se* are ineffective in altering the intellectual performance capability of the infant or young child. If, on the other hand, the intervention procedures are effective, then environment and changes in environment are a useful tool for altering intellectual performance capabilities. This would support a learning theorist's position in terms of the development of intelligence. The use of intervention then becomes highly relevant in discussing the issue of infant intelligence. Indeed, one might argue that this is one important method for getting at the effectiveness of environment on the infant's intellectual capacity. From an extreme interactionist point of view, the infant's intellectual capability is determined by the environment in which he exists. Thus, intervention programs should be crucial in determining whether or not this is the case.

We have already discussed the problems of measuring the effects of intervention by pointing out that infant intelligence cannot be considered some unitary construct measured by a single instrument. Moreover, it is necessary to match the evaluation of the intervention with the appropriate instrument. Thus, if one were affecting object permanence capacity in the young infant by such interventions as peek-a-boo games and showing the children how to find hidden objects, then the type of measurement should <u>not</u> be the Peabody Picture Vocabulary Test or some other verbal task, but rather a specific measure of sensorimotor capacity such as the Object Permanence Scales developed by Hunt and Uzgiris (1966) or Escalona and Corman (1967). Note that King and Seegmiller's (1971) results would argue against using the entire sensorimotor scales since they are not all necessarily related to object permanence. Thus, if we are to use intervention programs as a means of assessing whether infant's intellectual ability is fixed and unalterable by environment, then we must match the nature of the intervention procedure to the criterion of effectiveness. This, unfortunately, is rarely done.

Of even more importance is the notion that all children can benefit from the same kind of intervention procedure. This is a naive view, yet, unfortunately, widely held. Under this model every child in an intervention procedure must receive the same treatment, either for the sake of "scientific objectivity" or technical simplicity. Thus, every child is to watch a particular TV program or be instructed about a particular concept using a particular set of instructional material. The popularity of this model is surprising since both the educational experience in the classroom and more recently the educational psychology literature have increasingly realized that children need different types of intervention programs to arrive at the same goal, because children come into the intervention procedure with different kinds of experience. In the educational literature this issue refers to the aptitude-treatment interaction or subject-treatment interaction. In order to reach the same goal it is often necessary to apply different kinds of interventions (have various curricula), dependent upon the characteristics of the child. In a recent review, Berliner and Cahen (1973) make a strong argument for this position in educational programs and, more importantly, in the evaluation of their effectiveness. It is important to remember that it is the evaluation of environment effectiveness which is one way to consider the effectiveness of learning on the child's intellectual capacity.

An example of how we create difficulties by not considering the subject-treatment interaction in evaluating the effectiveness of the child's experience in the intervention program is necessary. Assume that we have 100 children in intervention Program A and that 10% of these children show increases in some measurement of the effectiveness of the intervention A. Ninety percent show no effect or even show some negative effect of intervention A. When we look at the data of the experimental group, averaged over all 100 children, we must conclude that intervention A was not a success. If we have 10 different intervention programs, each of them helping a different 10% of the experimental group, we would conclude that each of these programs failed to affect the measured capacity of these 100 children. Thus, intervention *per se* seems ineffective in influencing the child's intellectual capacities. In fact, this was not the case. All 10 intervention programs succeeded in affecting the child's intellectual capacity but did not do so for the group as a whole. Thus, across all 10 programs all 100 children's intellectual capacities, at least those that were designed to be affected by the intervention procedures, did show improvement. However, when we look at mean data we cannot locate any significant positive effect.

This example argues most powerfully for a subject-treatment interaction design both in terms of the nature of the program to be used and in terms of the evaluation of the effectiveness of that program.

It becomes then the function of the experimenter, curriculum developer, evaluator and finally theoretician to find what conditions each individual child needs to optimize his intellectual capacity. It is implicit in this assumption that (1) it is possible to find such conditions and that (2) having found such conditions we can come close to minimizing differences in intellectual capacity. Until such a program is initiated and until such a philosophy is undertaken, it is not fair to conclude that the intervention *per se* is ineffective or further that intellectual capacities cannot be affected by environmental change.

References

Bayley, N. Mental growth during the first three years: A developmental study of sixty-one children by repeated tests. Genetic Psychology Monographs, 1933, 14, 1-92.

Bayley, N. Comparisons of mental and motor test scores for ages 1-15 months by sex, birth order, race, geographical location and education of parents. Child Development, 1965, 36, 379-411.

Bayley, N. The Bayley Scales of Infant Development. New York: The Psychological Corporation, 1969.

Bayley, N. Development of mental abilities. In P. H. Mussen (Ed.), Carmichael's manual of child psychology. Vol. 1. New York: Wiley, 1970.

Berliner, D. C., & Cahen, L. S. Trait-treatment interactions and learning. In F. N. Kerlinger (Ed.), Review of research in education, 1973. Itasca, Ill.: F. E. Peacock Publishers, 1973, in press.

Burt, C., Jones, E., Miller, E., & Moodie, W. How the mind works. New York: Appleton-Century-Crofts, 1934.

Cole, M., Gay, J., Glick, J. A., & Sharp, D. W. The cultural context of learning and thinking. New York: Basic Books, 1971.

Escalona, S. K., & Corman, H. H. The validation of Piaget's hypotheses concerning the development of sensori-motor intelligence: Methodological issues. Paper presented at the Society for Research in Child Development meetings, 1967.

Guilford, J. P. Three faces of intellect. American Psychologist, 1959, 14, 469-479.

Hunt, J. McV., & Uzgiris, I. C. Scales of perceptual cognitive development. Unpublished manuscript, University of Illinois, 1966.

King, W., & Seegmiller, B. Cognitive development from 14 to 22 months of age in black, male, first-born infants assessed by the Bayley and Hunt-Uzgiris scales. Paper presented at the Society for Research in Child Development meetings, Minneapolis, 1971.

Lewis, M. State as an infant-environment interaction: An analysis of mother-infant behavior as a function of sex. Merrill-Palmer Quarterly, 1972, 18, 95-121.

Lewis, M., & Lee-Painter, S. An infant's interaction with its social world: The origin of meaning. Paper presented at Canadian Psychological Association meetings, Symposium on Parent-Child Observation Studies and Their Problems, Montreal, June 1972.

Lewis, M., & McGurk, H. The evaluation of infant intelligence: Infant intelligence scores--true or false? Research Bulletin 72-32. Princeton, N. J.: Educational Testing Service, 1972. Science, in press.

McCall, R. B., Hogarty, P. S., & Hurlburt, N. Transitions in infant sensori-motor development and the prediction of childhood IQ. American Psychologist, 1972, 27(8), 728-748.

Stott, L. H., & Ball, R. S. Infant and preschool mental tests: Review and evaluation. Monographs of the Society for Research in Child Development, 1965, 30, Serial No. 101.

Thomas, H. Psychological assessment instruments for use with human infants. Merrill-Palmer Quarterly, 1970, 16, 179-224.

Footnote

[1]This research was supported by a grant from the National Science Foundation, GB 8590.